EXAMINATION OF POTENTIAL BIASES IN AIR TEMPERATURE CAUSED BY POOR STATION LOCATIONS

BY THOMAS C. PETERSON

Analysis of a small subset of U.S. Historical Climatology Network data does not find a time-dependent bias caused by current poor station siting.

Changing the instrumentation, location, or observing practices at in situ weather stations introduces nonclimatic biases into the data. During the last few decades, a great deal of effort has gone into developing methods to adjust in situ station temperature time series to account for these artificial changes or inhomogeneities in the climate record. Reviews of homogeneity testing and adjustment techniques indicate that many approaches successfully remove artificial discontinuities from the time series caused by a wide variety of types of changes (Aguilar et al. 2003; Peterson et al. 1998). But, can these approaches compensate for problems caused by poor siting and particularly changes to siting? Davey and Pielke (2005; hereafter Davey and Pielke) performed an excellent analysis of the microclimate exposures of weather-observing stations in eastern Colorado that found that the siting of many stations does not conform to National Weather Service or World Meteorological Organization siting standards. Indeed, they concluded that sites with good temperature-exposure characteristics were in the minority. They also expressed concern that the poor siting could be causing a bias in the temperature record, but, as noted by Vose et al. (2005), did not actually analyze the data to determine if poor siting resulted in spurious trends or not.

AFFILIATIONS: PETERSON—NOAA's National Climatic Data Center, Asheville, North Carolina
CORRESPONDING AUTHOR: Thomas C. Peterson, National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801
E-mail: thomas.c.peterson@noaa.gov

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-87-8-1073

In final form 20 April 2006

Essentially there are two competing hypotheses about the effects of poor siting that yield very different predictions. The first hypothesis is that homogeneity-adjustment methodologies would account for changes to locations with poor siting. If the homogeneity adjustments are appropriately accounting for all artificial changes at the stations, then an adjusted temperature time series from the poorly sited stations should be very similar to the time series from the stations with good siting. The trends from the poorly sited stations may be a little higher or a little lower, but they should still be about the same. This hypothesis would, of course, also hold if poor siting did not cause a bias in the original data and the homogenization did not introduce any biases. The second hypothesis is that poor current station siting produces an artificial bias in the temperature record that is not being addressed by homogeneity adjustments. While Davey and Pielke suggested that poor siting–induced bias could be positive or negative, the underlying concern about the effects of potential siting biases is whether a significant portion of the recent warming indicated by the U.S. and global temperature record could be due to this bias rather than climate change. Therefore, the second hypothesis predicts that homogeneity-adjusted temperature trends at the poorly sited station would be significantly different than the temperature trends at the stations with good siting, and that these differences would most likely be that the poorly sited stations are warming relative to nearby stations with good siting.

The analysis presented in this paper builds on the work of Davey and Pielke by performing analyses on data from some of the stations that they identified as having good and poor siting, with the intent of testing these two hypotheses.

STATIONS, DATA, AND HOMOGENEITY ADJUSTMENTS. Of the stations inspected by Davey and Pielke, only those stations that make up part of the U.S. Historical Climatology Network (USHCN; Easterling et al. 1996) have both the original (raw unadjusted) data and data that have been adjusted to account for inhomogeneities. While Davey and Pielke report that 10 USHCN stations were inspected, only 8 of these stations were described in detail. Two stations were listed as having good exposure: Trinidad and Cheyenne Wells. Two stations were described as having questionable site exposure or a mixture of conditions, such as being well ventilated but near a gravel road, and were not used in this analysis. Four stations were listed as having poor exposure: Eads, Lamar, Las Animas, and Holly. These sites had multiple problems, with the most dominant one being that they were sited too close to obstructions, such as houses.

The locations of these stations, excluding Holly for reasons described later, are given in Table 1. The mean annual maximum and minimum temperature at each of these stations is also provided in Table 1. For many purposes for which station data are used, the actual observed temperatures are used directly. However, to examine the change through time, temperature time series are typically converted into anomaly time series by subtracting out the mean temperature from a base period, such as 1971–2000 (e.g., Jones and Moberg 2003). Unlike actual temperatures, station anomaly values can be averaged together without adversely impacting the time series when a particularly warm or cold location has some observations missing, because temperature anomalies are much more geographically coherent than actual temperatures.

Because the analysis presented here only examines changes in temperature over time, all of the results will be using anomaly time series. Examination of poor siting–induced biases by comparing the absolute value of temperature observations at neighboring stations must consider the confounding effects of general site topography (e.g., exposure to cold air drainage), observing practices, instrumentation, latitude, and elevation. For example, it is perhaps not surprising that the station with the warmest temperature in Table 1 has the lowest elevation. The time of day the thermometers are read can make as much as a 1.4°C difference in mean annual maximum or minimum temperature in this part of the country (Karl et al. 1986). Furthermore, Gallo (2005) found microclimate-related differences exceeding 0.5°C

Table 1. Stations used in this analysis, their location, and their mean annual 1969–2004 unadjusted maximum and minimum temperature for years when all 12 months of data were present.

Name	ID no.	Lat	Lon	Elevation (m)	Mean annual maximum temperature (°C)	Mean annual minimum temperature (°C)
Trinidad	058429	37°11'N	104°29'W	1839	19.5	2.9
Cheyenne Wells	051564	38°49'N	102°22'W	1295	19.2	2.9
Eads	052446	38°29'N	102°47'W	1285	19.8	2.6
Lamar	054770	38°06'N	102°38'W	1267	20.5	3.3
Las Animas	054834	38°03'N	102°07'W	1033	21.6	3.2

in pairs of stations, differences that could not be explained by either latitude, elevation, instrumentation, observing practices, or quality of the siting. As examination of Table 1 reveals, the mean absolute differences in temperature at the stations with good and poor siting varies within the range one might expect from these factors.

The goal of this work is not to evaluate absolute biases due to poor siting, but rather to evaluate any potential time-dependent aspect of siting-induced biases. These time-dependent biases can either be caused by changes to poor siting, or they may be due to different siting-induced responses to climate change. For example, a station located over an impermeable surface will not experience the same microclimate-induced changes in temperature caused by changes in local latent versus sensible heat release as that which a station located over grass would likely experience during a wet (dry) spell when adequate soil moisture is (not) available for latent heat release by the grass. If precipitation changed over time, this could theoretically lead to a bias in observed temperature at a poorly sited station compared to a station with good siting. On the other hand, both stations would experience exactly the same number of cold fronts and other synoptic-scale weather events whose influences may swamp any impact from changing microscale conditions.

The current USHCN adjustment methodology is based on metadata. If a station history file indicates that a change, for example, in instrumentation or station location, took place, the historical record is adjusted up or down in an attempt to make it equivalent to what would have been observed by the current instrumentation at the current observing location with an observing time of midnight. Table 2 shows the reasons for and dates of the homogeneity adjustments made at these five stations (two with good siting, three with poor siting) during the last four decades, along with the value of the adjustments. The only exception is that, because the focus of this analysis is only on change over time, the value of the

Table 2. Dates, reasons, and values of the homogeneity adjustments applied to the station time series. Every change in instrumentation in this table was a change from a liquid-in-glass thermometer in a CRS to the electronic MMTS. Major time of observations (TOB) changes are when, for example, a morning reader becomes an afternoon reader. Minor changes are when, for example, a morning observer stays a morning observer but the time of observation changes from 0900 to 0700. In addition to the magnitude and the sign of the adjustment (°C), the influence that the two stations with good siting, Trinidad and Cheyenne Wells, had in determining the adjustment is given in (%).

Station	Year	Type	Tmax		Tmin		Tmean	
			Adjustment (°C)	%	Adjustment (°C)	%	Adjustment (°C)	%
Trinidad	None		0	0	0	0	0	0
Cheyenne Wells	1987	Minor TOB	+0.62	0	+0.05	0	+0.34	0
	1981	Minor TOB	−0.45	0	−0.03	0	−0.24	0
Eads	1993	Minor TOB	0.00	0	−0.32	0	−0.16	0
	1987	Major TOB	−0.21	0	−0.92	0	−0.57	0
	1986	Instrumentation	−0.38	0	+0.28	0	−0.02	0
	1982	Major TOB	−1.12	0	−0.05	0	−0.58	0
	1981	Relocation	+0.36	42	+1.52	26	+0.89	37
Lamar	1992	Minor TOB	+.01	0	−0.31	0	−0.16	0
	1991	Relocation	−0.68	28	+1.12	22	+0.21	22
	1989	Major TOB	−0.12	0	−0.37	0	−0.25	0
	1988	Instrumentation	−0.38	0	+0.28	0	−0.03	0
	1986	Relocation	+0.95	28	−2.26	23	−0.83	32
	1979	Minor TOB	−0.6	0	−0.21	0	−0.13	0
	1978	Relocation	−0.13	24	+0.24	30	0.00	31
Las Animas	1989	Major TOB	−0.88	0	−0.57	0	−0.72	0
	1986	Instrumentation	−0.37	0	+0.28	0	−0.04	0

adjustment to the current data to make them equivalent to a midnight observer is not included. Each of the homogeneity adjustments listed are added to all of the original raw data prior to that discontinuity to make the homogeneity-adjusted dataset. One of the two stations with good siting had no homogeneity adjustments during this period, and the other only had two relatively minor changes in the time of observation that primarily impacted maximum temperature, with the second change six years later largely offsetting the first change. No change in adjustments of any kind was made to data from any of the stations after 1993.

Unfortunately though, some station changes are not documented in the station history file. A new adjustment methodology for the USHCN that also uses statistical techniques to find undocumented changes is in the evaluation phase (Williams and Menne 2005). The preliminary results of this technique indicated that one of the six USHCN stations identified by Davey and Pielke had an undocumented change—Holly in 1996. Therefore, Holly, one of the poorly sited stations, was not included in the analysis.

Three different homogeneity adjustments were applied to the data. The adjustment that often makes the biggest difference is for changes in the time of observation (Karl et al. 1986). The formula used for calculating the appropriate adjustment varies with station location, month, and, of course, the time the observations were made. Recently reevaluated by Vose et al. (2003), this adjustment was found to be quite accurate in the United States as a whole. Another adjustment accounts for the change in instrumentation from liquid-in-glass thermometers in Cotton Region Shelters (CRS) to the electronic maximum–minimum temperature system (MMTS; Quayle et al. 1991). While the adjustment is seasonally varying, all stations in the USHCN undergoing this change in instrumentation get the same adjustment factor applied to their data. In Table 2 the instrumentation adjustment for mean temperature varies slightly between the stations because of the way rounding was addressed in the software making the mean temperature adjustments.

The last adjustment is to account for station moves (Karl and Williams, 1987). This is the only homogeneity adjustment that determines the appropriate adjustment by comparing the station data with other nearby stations. For example, in response to metadata indicating that the Lamar station moved in 1978, station histories were examined to find nearby stations with no documented changes in the five or more years on either side of that date to use as reference stations. Seasonal Lamar time series were then compared with the time series from the reference stations in a statistical procedure to determine if the change resulted in a statistically significant discontinuity, and, if so, an adjustment value would be applied to the Lamar temperature time series.

This adjustment means that the two stations with good siting may have directly contributed to the homogeneity adjustments of the stations with poor siting. To determine how much of an effect this might be, intermediate output files were examined that document the contribution of each station to each adjustment. It turns out that Trinidad data were not used in any of the adjustments, but Cheyenne Wells data were. Table 2 also indicates the mean weight of Cheyenne Wells used in determining the magnitude of the adjustments.

It turns out that the influence of Cheyenne Wells on the total adjustments to the three stations with poor siting is quite minor. Table 3 provides a summary of the total magnitude of all of the homogeneity adjustments applied to Eads, Larmar, and Las Animas on a per-station-average basis. Time-of-observation adjustments impacted maximum, minimum, and mean temperature in about the same magnitude. The adjustment to account for changes in instrumentation from CRS to MMTS mainly impacted maximum and minimum temperature, with only a small effect on mean temperature. Station relocations, as one might expect, had a much greater impact on minimum temperature than on maximum temperature because small changes in exposure or topography do not have as much of an effect on observed temperature in a well-mixed boundary layer as they can in a stably stratified environment. The magnitude of the relocation adjustments attributable to Cheyenne Wells is also provided. As indicated in Table 3, these amount to only 11%–14% of the total adjustments applied to the data from the three stations with poor siting. Therefore, while the homogeneity-adjusted data from the stations with poor siting are not totally independent of the data from the stations with good siting, the dependence is small.

There is one other modification of the data. When data are missing or when there is a very short period for which reliable homogeneity adjustments are not possible, data for USHCN stations are filled in with interpolated values. The start of the period examined, 1969, was chosen to avoid a period with interpolated data at Eads. Of the five stations, only Lamar had no interpolated values. Approximately 2% of the station months used in this analysis had interpolated values; three-quarters of these were in the 1980s, and the majority were at Eads.

TABLE 3: Summary of the magnitude of the homogeneity adjustments. Absolute value of the time of observation, instrumentation, and relocation adjustments (°C) applied to the Eads, Lamar, and Las Animas time series on an average-per-station basis. The sum of the absolute value of these three adjustments is provided in the row labeled total adjustments. The magnitude of the portion of the relocation adjustments attributable to Trinidad and Cheyenne Wells (the magnitude of the relocation adjustment times the fraction attributable to these stations as shown in Table 2) is provided (°C). The bottom row is the percent of the total absolute value of the adjustments that can be attributed to the two stations with good siting.

Type	Tmax	Tmin	Tmean
Time of observation	0.80°C	0.92°C	0.86°C
Instrumentation	0.38°C	0.28°C	0.03°C
Relocation	0.71°C	1.71°C	0.64°C
Total adjustments	1.88°C	2.91°C	1.53°C
The magnitude of the portion of relocation adjustments attributable to Trinidad and Cheyenne Wells	0.21°C	0.41°C	0.21°C
Percent of total adjustments impacted by Trinidad and Cheyenne Wells	11%	14%	14%

RESULTS. The average anomaly (from the 1971–2000 base period) time series for 1969 through 2004 from the two stations with good exposure are shown in Fig. 1 for both the unadjusted (dashed) and homogeneity-adjusted (solid) versions of maximum (red) and minimum (blue) temperature. Cheyenne Wells did have small adjustments, much of which were counteracted by another adjustment seven years later, and therefore the unadjusted and adjusted temperature for the stations with good siting are slightly different.

The adjusted maximum, minimum, and mean time series from the two stations with good siting are a priori likely to be quite representative of the region because a) the data come from stations with excellent siting (Davey and Pielke), b) the metadata indicate that they needed very little homogeneity adjustments, c) the homogeneity adjustment they did need was recently reevaluated and found to be quite accurate for the United States as a whole (Vose et al. 2003), and d) the statistical homogeneity tests indicated no non-metadata-reported inhomogeneities (Williams and Menne 2005). Therefore, the following figures will show comparisons between these reliable time series and time series from the three stations with poor siting.

Figures 2 and 3 show the maximum and minimum temperature time series. In both figures the black time series is the adjusted time series from the stations with good siting, with the colored lines representing the unadjusted (dashed) and homogeneity-adjusted (solid) time series derived from the data from the three poorly sited stations. In both cases the homogeneity adjustments changed the trend in the time series considerably, even changing the sign in the case of maximum temperature. Also, in both cases, the adjustments made the trends very similar to the trends from the two stations with good sighting, but indicate slightly less warming than the stations with

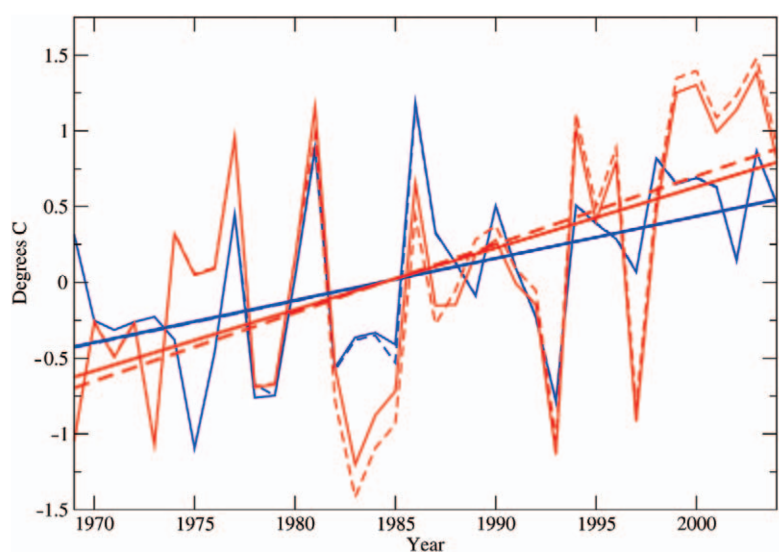

FIG. 1. Mean anomaly of the two stations with good siting's annual maximum temperature (red) and minimum temperature (blue) and their linear trends. Original unadjusted (dashed) and adjusted (solid) data. As indicated in Table 2, the homogeneity adjustments to these data were small.

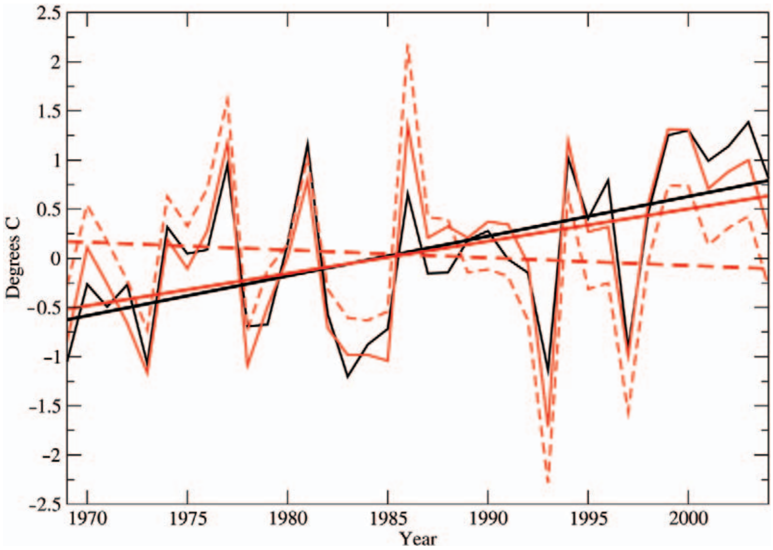

FIG. 2. Mean annual anomalies of maximum temperature data and their linear regressions from the two stations with good siting (black) homogeneity-adjusted data and the three stations with poor siting for both original unadjusted (red dash) and adjusted (red solid) data. The homogeneity adjustments applied to the stations with poor siting makes their trend very similar to the trend at the stations with good siting. Although the adjustments are designed to make the historical anomalies consistent with the most recent data, the adjustments applied within the 1971–2000 base period used to calculate the anomalies can create an offset in the most recent anomaly data.

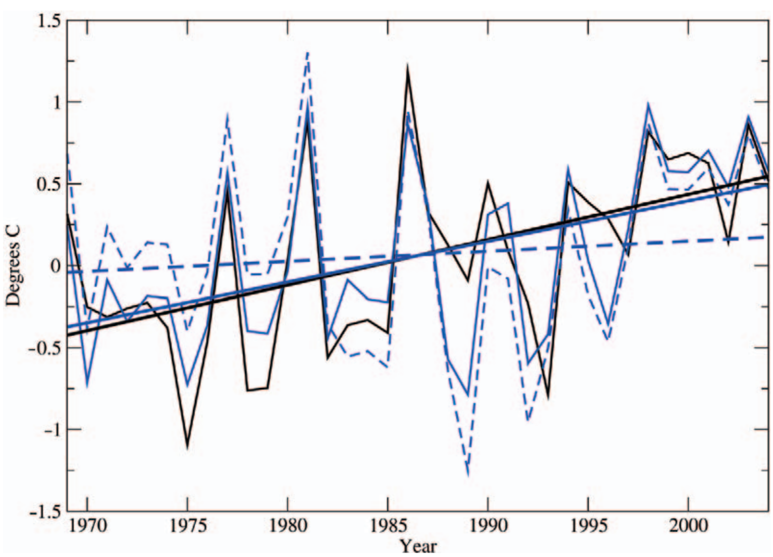

FIG. 3. Mean annual anomalies of minimum temperature from the homogeneity-adjusted data and their linear regressions from the two stations with good siting (black) and the three stations with poor siting for both original unadjusted (blue dash) and adjusted (blue solid) data. Again, the homogeneity adjustments applied to the stations with poor siting make their trend very similar to the trend at the stations with good siting.

good siting. The slight effect of filling in for missing data can be seen in the offset between the adjusted and the raw maximum temperature data after 1993 (shown in Fig. 2). The offset is 0.57°C for 1993 and for each year from 1996 through 2004, but it is 0.02°C less in 1994 and 0.01°C more in 1995 due to the filling in of five months of data in 1994 and four months of data in 1995 in the Eads time series.

Perhaps the most important analysis is that of mean temperature, because the mean temperatures from these stations contribute to the global land and ocean temperature time series. Figure 4 shows the homogeneity-adjusted time series from both the stations with good siting (black) and the stations with poor siting (magenta), along with their regression lines. It is striking that not only do the two time series have very similar trends, with the poorly sited stations showing slightly less warming, but the year-to-year variability is quite similar as well. It should be noted that the dataset with slightly more warming than the other could be different if slightly different years were analyzed. Also, if the analysis had included the incompletely homogenized data from Holly, the results would have indicated somewhat less warming at the stations with poor siting. In all three cases of maximum, minimum, and mean temperature, the adjustments brought the quite different time series from the poorly sited stations into close agreement with the homogeneous time series from the stations with good siting.

To further examine the differences between the time series, the data in the first and last third of the time series were compared. The homogeneity-adjusted data were differenced by subtracting the time series from the poorly sited stations from the time series of the stations with good siting. Figure 5 shows a box-and-whisker representation of this difference series for these two periods of time. The null hypothesis that the difference in temperature anomalies in these two periods was not significantly different was tested using a multiresponse permutation test (MRPP; Mielke 1991). The difference between the first 12 and last 12 years was not found to be significant at the 5% level for any of the three datasets.

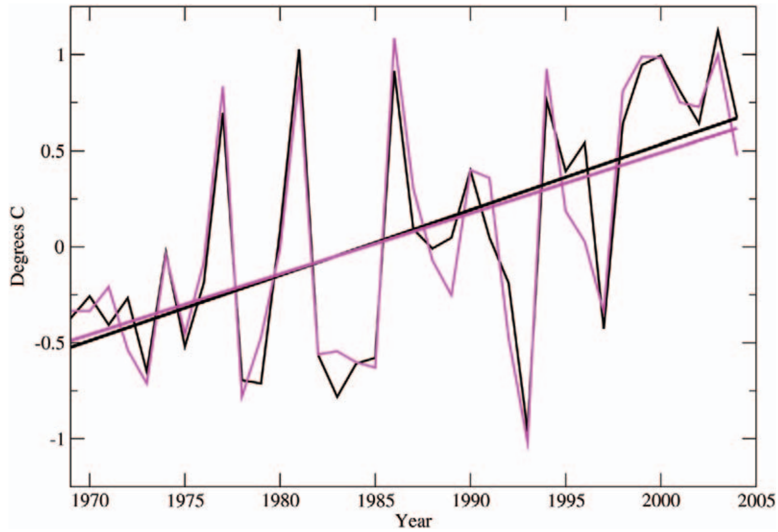

FIG. 4. Annual homogeneity-adjusted mean temperature anomaly time series and their linear regressions from the two stations with good siting (black) and the three stations with poor siting (magenta). Not only are the trends similar, but so are their year-to-year variations.

CONCLUSIONS. Classically, science progresses by developing hypotheses that lead to predictions that can be evaluated by comparison with physical reality. Each successful prediction adds to the weight of evidence supporting the theory, and any unsuccessful prediction demonstrates that the theory is imperfect and requires improvement or abandonment. Because the number of stations evaluated in this study is quite limited, the results cannot be definitive, but they can supply some evidence in support or rejection of a hypothesis. The results presented here clearly support the theory that, if poor siting causes a bias, homogeneity adjustments account for the biases and contradict the hypothesis that poor current siting causes a warm bias or even any bias in the homogeneity-adjusted U.S. temperature change record.

The homogeneity-adjusted time series from the two stations with good siting are a priori likely to be representative of the climate trends and variability of the region, because their data were nearly homogeneous to begin with and a thoroughly evaluated homogeneity adjustment was used to account for the temporary, relatively minor, change in time of observation at one of these stations. Furthermore, the close agreement with the homogeneity-adjusted data from the stations with poor siting make a strong a posteriori case that data from the two stations with good siting are indeed representative of the climate of the area. Slight unrepresentativeness may still arise, however, because climatic changes and variations may differ slightly with altitude, latitude, longitude, and natural land surface. The adjustments at the stations with poor siting were, for the most part, independent of the well-sited stations, but changed their composite time series from being very different to agreeing very well with the time series from the well-sited stations, indicating that the homogeneity adjustments applied to the data from the poorly sited stations compensated for bias-producing changes. This result also suggests that the wider set of stations, after adjustment of the data from poorly sited stations, is truly representative of the climate trends and variability of the region.

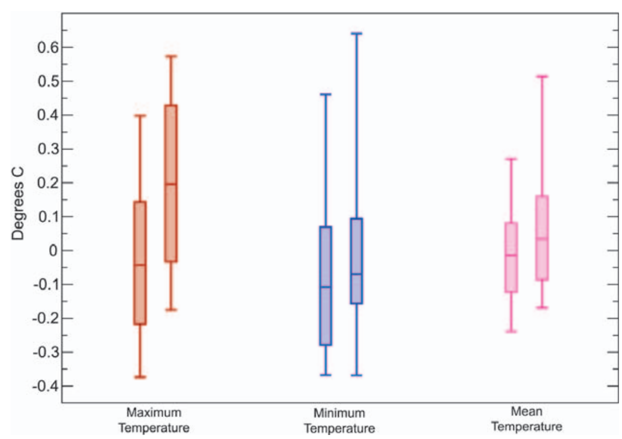

FIG. 5. Box-and-whiskers plots of the (left) first and (right) last 12 years of the difference between the homogeneity-adjusted time series from the two stations with good siting and (minus) the homogeneity-adjusted time series from the three stations with poor siting for maximum, minimum, and mean temperature. The center-most line in the boxes is the median value, the edge of the boxes are the 25th and 75th percentiles, and the end of the whiskers are the extreme values. None of the differences between the first and the last 12 years were significant at the 5% level.

It should be noted, though, that new techniques are needed, because the results would not have been as good if the station with the statistically detected inhomogeneity that was unsupported by available metadata was included in the analysis. But, in the end, the similarity between the homogeneity-adjusted time series from the good and poorly sited stations supports the view that even stations that do not, upon visual inspection, appear to be spatially representative can, with proper homogeneity adjustments, produce time series that are indeed representative of the climate variability and change in the region.

Because weather data have a myriad of different uses, the results of an analysis related to one particular use cannot justify station siting practices that do not meet national and international standards. Data that do not meet quality standards necessary for particular analyses have caused numerous scientists at the National Climatic Data Center and elsewhere around the world to spend years, and indeed decades, developing techniques to improve the fidelity of in situ data for their particular applications. This analysis takes the opportunity afforded by the work of Davey and Pielke to evaluate not only the effects of poor station siting, but also the homogeneity-adjustment techniques painstakingly developed over many years at the National Climatic Data Center. The results indicate that the work was not done in vain: the homogeneity adjustments did an excellent job of accounting for time-dependent biases at the stations examined and the homogeneity-adjusted data do not indicate any time-dependent bias caused by current poor station siting.

ACKNOWLEDGMENTS. This work is supported by NOAA Climate Program Office Climate Change Data and Detection Program Element and a Department of Energy Interagency Agreement. The author appreciates the help by Claude Williams on adjustment details and useful discussions with Thomas Karl and Jack Pfitsch. Also, the comments by all five of the *BAMS* reviewers helped to improve the article.

REFERENCES

Aguilar, E., I. Auer, M. Brunet, T. C. Peterson, and J. Wieringa, 2003: Guidelines on climate metadata and homogenization. World Meteorological Organization WCDMP-No. 53, WMO-TD No. 1186, 55 pp.

Davey, C. A., and R. A. Pielke Sr., 2005: Microclimate exposures of surface-based weather stations. *Bull. Amer. Meteor. Soc.,* **86,** 497–504.

Easterling, D. R., T. R. Karl, E. H. Mason, P. Y. Hughes, D. P. Bowman, R. C. Daniels, and T. A. Boden, 1996: United States Historical Climatology Network (U.S. HCN) monthly temperature and precipitation data. Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory Publication 4500, 83 pp. + appendixes.

Gallo, K. P., 2005: Evaluation of temperatures differences for paired stations of the U.S. Climate Reference Network. *J. Climate,* **18,** 1631–1638.

Jones, P. D., and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate,* **16,** 206–223.

Karl, T. R., and C. N. Williams Jr., 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Meteor.,* **26,** 1744–1763.

——, ——, P. J. Young, and W. M. Wendland, 1986: A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States. *J. Climate Appl. Meteor.,* **25,** 145–160.

Mielke, P. W., 1991: The application of multivariate permutation methods based on distance functions in the earth sciences. *Earth-Sci. Rev.,* **31,** 55–71.

Peterson, T. C., and Coauthors, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.,* **18,** 1493–1517.

Quayle, R. G., D. R. Easterling, T. R. Karl, and P. Y. Hughes, 1991: Effects of recent thermometer changes in the Cooperative station network. *Bull. Amer. Meteor. Soc.,* **72,** 1718–1724.

Vose, R., S., C. N. Williams, T. C. Peterson, T. R. Karl, and D. R. Easterling, 2003: An evaluation of the time of observation bias adjustment in the U.S. historical climate network. *Geophys. Res. Lett.,* **30,** 2046, doi:10.1029/2003GL018111.

——, D. R. Easterling, T. R. Karl, and M. Helfert, 2005: Comments on "Microclimate exposures of surface-based weather stations." *Bull. Amer. Meteor. Soc.,* **86,** 504–506.

Williams, C. N., Jr., and M J. Menne, 2005: The U.S. Historical Climate Network, Version 2. *15th Conf. on Applied Climatology,* Savannah, GA, CD-ROM, 1.1.